전처리의 고속화에 기반한 문자 인식 시스템

공용해[†]

요 약

다량의 문자가 온라인으로 실시간 입력되는 문자인식시스템에서는 영상의 전처리가 매우 신속하게 이루어져야 한다. 또한 크기가 작거나 잡음을 포함하는 문자 영상에 대한 기하학적 정규화나 세선화 등의 영상 변환은 원영상의 정보 유실이나 왜곡을 유발하므로 최소화되어야 한다. 이에 따라 본 연구는 원영상을 변환하지 않은 상태에서 신속하게 인식에 효과적인 특징을 구하고자 하였다. 이를 위해 인식의 관점에서 의미를 가지는 윤곽화소를 정의하고, 다양한 영상의 변화를 수용하는 효과적인 특징을 윤곽화소로부터 구하였다. 필기체 자모음과 자동차 번호판 숫자 인식 실험 결과, 제안된 방법이 기존의 방법보다 인식에 보다 효과적이었으며 한번의 영상 주사로 전처리를 완료하므로 소요 시간을 크게 줄일 수 있었다.

Character Recognition System using Fast Preprocessing Method

Yong Hae Kong[†]

ABSTRACT

A character recognition system, where a large amount of character images arrive continuously in real time, must preprocess character images very quickly. Moreover, information loss due to image transformations such as geometric normalization and thinning needs to be minimized especially when character images are small and noisy. Therefore, we suggest a prompt and effective feature extraction method without transforming original images. For this, boundary pixels are defined in terms of the degree in classification, and those boundary pixels are considered selectively in extracting features. The proposed method is tested by a handwritten character recognition and a car plate number recognition. The experiments show that the proposed method is effective in recognition compared to conventional methods. And an overall reduction of execution time is achieved by completing all the required processing by a single image scan.

1. 서 론

멀티미디어 응용 분야 중에서 시각 정보의 자동화 처리는 그 실질적 필요성이 매우 크다 할 수 있으며, 특히 인쇄되거나 필기된 문서 정보의 오프라인 문자인식시스템은 문서인식, 전표인식, 우편번호인식, 차량번호인식 등 그 요구 분야가 광범위하고 실용성이 크므로 지속적으로 연구되고 있다[1,2,3]. 문자인식시스템은 문자 자체의 형태 변화뿐만 아니라 크기, 위치 등의 기하학적 변화와 다양한 획 폭의 변화를 수용할 수 있도록 설계되어야 한다[4]. 문자 영상의 획의 폭을 일정하게 하는 방법으로 세선화(thinning) 알고리즘이 많이 이용되고 있으나, 세선화는 잡음이나 세부 형태의 변화에 민감하므로 세선화 후의 골격(skeleton)이 문자 영상의 원래 형태를 적절히 유지하지 못할 수 있다[5,6]. 산업 현장에서 취득되는 문자 영상은 크기가 작고 잡음을 포함하는 경우가 많으므로 문자 영상에 대한 기하학적 정규화나 세선화 등의 영상 변환은 원영상이 왜곡되어 인식에 부정적인 효과를 유발할 수 있다. 따라서 원영상을 변환하지 않은 상태에서 인식에 유리한 효과적인 특징을

이 논문은 1997년 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

[†] 순천향대학교 컴퓨터학부 부교수

구하는 것이 바람직하다.

대부분의 문자인식시스템은 문자 영상의 전처리와 신경회로망의 학습으로 구분될 수 있다. 신경회로망의 학습은 고속의 컴퓨터나 분산, 병렬 컴퓨팅 환경에서 비교적 오랜 시간 오프라인으로 이루어질 수 있으며, 최종 학습된 상태만을 비교적 저속인 일반 컴퓨터에 이식시킬 수 있다. 그러나 인식에 필요한 제반 문자 영상의 전처리는 다량의 데이터를 제한된 시간 내에 처리해야 하므로 매우 신속하게 이루어져야 한다.

본 연구는 오프라인 문자인식시스템의 구현에 있어 다음의 두 가지 사항을 고려하였다. 첫째, 인식을 위한 신경회로망의 학습에는 상대적으로 시간적 제약을 크게 두지 않은 반면에 인식에 필요한 제반 전처리 과정을 신속하게 수행하고자 하였다. 둘째, 원 영상에 포함된 모든 정보를 인식에 최대한 이용하기 위하여, 크기, 위치의 정규화나 세선화 목적으로 영상을 변환시키지 않고 인식에 적합한 특징을 원영상에서 직접 구하고자 하였다. 이를 위하여 인식 측면에서 의미를 가지는 윤곽화소를 선택하여 방향 특징을 구하였으며, 문자 영상의 위치, 크기, 획의 폭 변화를 수용하기 위하여 윤곽에서 추출된 방향 성분을 영상 축에 투영한 비례값을 사용하였다. 전처리에 소요되는 시간을 줄이고자 상기의 과정을 한 번의 영상 주사(scan)로 이루어지도록 하였다. 제안된 방법이 인식에 효과적인지 실험하기 위하여, 필기체 자모음 인식과 자동차 번호판 숫자 인식을 실험하였다. 또한 비교 실험으로서 크기와 위치 정규화, 세선화를 모두 수행하는 특징 추출 방법과 정규화 센트랄모우먼트(normalized central moment) 방법을 동일 문자 집단에 적용하여 인식률과 인식에 소요되는 시간을 비교하였다. 신경회로망의 학습에 있어서는 유전자 알고리즘을 신경회로망의 가중치 초기화 과정에 사용하여 수렴 효과를 조사하였다[7].

2. 문자 영상의 정규화, 특징 추출

스캐너나 카메라 등에 의해 문자 영상이 취득되면 먼저 잡음 제거와 이진화 후에 인식에 적합한 특징을 추출하게 된다. 문자 영상은 문자 자체가 가지는 다양한 형태 변화뿐만 아니라 크기, 위치, 회전에 의한 기하학적 변화를 포함하므로, 기하학적 변화를 특정 위치나 규모로 획일화시키는 과정이 인식에 앞서 선행된다. 문자 영상의 기하학적 변화가 보정된 이후에 문자의 형태를 효과적으로 대표하는 특징을 구한다. 간단하면서도 많이 사용되는 문자 영상의 특징으로 망 특징이 사용된다. 망 특징은 문자 영상을 그림 1(a)와 같이 일정 등분 구간으로 나누어 구간 내의 영상 화소수를 누계한 특징을 사용한다. 큰 특징 벡터는 패턴분류기의 입력에 부적절하므로 문자 영상의 전역적인 형태를 작은 특징 벡터로 나타내고자 할 경우 망 특징이 유리하다. 그림 1(b)와 그림 1(c)에서 보이듯이 문자 영상을 수평 방향과 수직 방향으로 일정 간격으로 주사하면서 영상과 배경이 교차하는 수를 세는 교차점수 특징과 교차하는 거리를 이용하는 교차거리 특징은 적절히 정규화된 경우 효과적일 수 있다. 그러나 망, 교차점수, 교차거리 특징은 문자 영상의 크기, 위치, 회전의 변화에 매우 민감할 뿐만 아니라 잡음이 더해진 경우에는 사용 시 제약이 따른다[8].

(a) 망 특징 (b) 교차점수 특징 (c) 교차거리 특징

그림 1. 망, 교차점수, 교차거리 특징

문자 영상의 경우 영상 내의 획 성분의 방향과 크기의 조합이 문자 형태를 특징 한다고 볼 수 있다. 획의 방향 성분을 알아내기 위해 보편적으로 마스크 연산자를 사용한다. 마스크 연산자로는 Sobel 마스크, 4방향 마스크(그림 2), Prewitt 마스크, Kirsch 마스크(그림 3) 등이 있다. 이러한 방향 마스크 연산 결과는 상기한 압축 특징과 함께 사용되기도 한다[9]. 이 외에 기하학적인 영상에 논리 연산을 가하는 모폴로지를 이용하는 방법과 영상을 수평축과 수직

0	0	0
1	1	1
0	0	0

수평 방향

0	1	0
0	1	0
0	1	0

수직 방향

0	0	1
0	1	0
1	0	0

사선 방향

1	0	0
0	1	0
0	0	1

역사선 방향

그림 2. 방향 마스크

5	5	5	-3	-3	-3	-3	-3	5	5	-3	-3
-3	0	-3	-3	0	-3	-3	0	5	5	0	-3
-3	-3	-3	5	5	5	-3	-3	5	5	-3	-3

수평 방향　　　　　　수직 방향

-3	-3	-3	5	5	-3	-3	5	5	-3	-3	-3
-3	0	5	5	0	-3	-3	0	5	5	0	-3
-3	5	5	-3	-3	-3	-3	-3	-3	5	5	-3

사선 방향　　　　　　역사선 방향

그림 3. Kirsch 마스크

축에 투영한 결과를 이용하는 방법도 사용되고 있다 [10].

이상 열거한 방법들은 자체적으로 기하학적인 변화를 수용하지 못하므로 크기, 위치, 회전 변화를 보상하는 전처리 과정을 필요로 한다. 영상의 기하학적 변화의 일부나 모두를 수용하는 특징으로 허프 변환(hough transform), 로그 폴라 변환(log polar transform), 퓨리어 디스크립터(fourier descriptor), 모우먼트 인베어리언트(moment invariant) 등이 사용될 수 있다. 이러한 방법들은 양자화 오류와 샘플링 오류에 의해 이진화된 작은 문자 영상에 적용하는 데 있어서 제한점으로 작용할 수 있다[11].

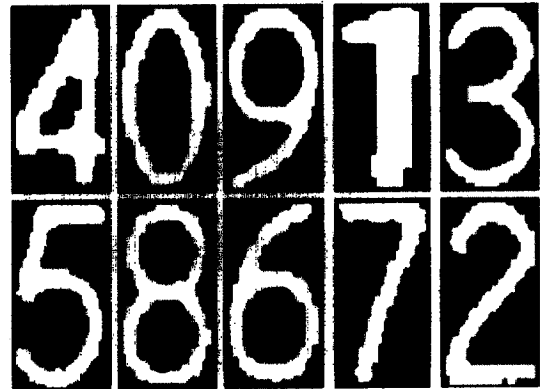
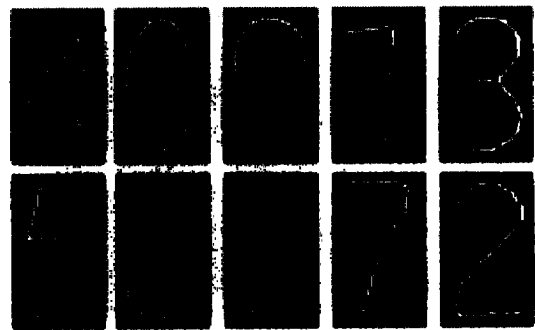
문자 영상은 획(stroke)의 두께가 일정하지 않는데 이러한 현상은 문자 영상이 인쇄체이거나 필기체이거나 관계없이 발생한다. 인쇄체인 경우 폰트나 프린터에 따라 다양한 획의 폭이 출력된다. 필기체인 경우 필기 스타일이나 필기구의 특성에 따라 역시 다양한 획의 폭이 결과된다. 뿐만 아니라 입력 시점에서의 입력 장치, 조명등의 환경과 이진화 시의 이진화 임계값(threshold)도 문자 영상의 획의 폭 변화에 지대한 영향을 준다. 결과적으로 문자 영상은 매우 다양한 폭을 가지는 획으로 구성된다고 할 수 있으며 획의 폭에 무관한 특징 추출은 인식에 있어 중요하다. 대부분 획 폭의 불규칙성을 최소화하기 위해 세선화 알고리즘을 사용하는데 세선화란 두께를 가진 획의 바깥쪽부터 화소를 하나씩 제거해서 획의 폭을 연결성을 유지하는 최소 폭인 하나의 화소로 만드는 과정이다. 세선화된 문자 영상은 모두 한 화소의 폭을 가지는 획으로 구성되어 문자 형태를 유지하는 골격(skeleton)으로 표현되게 된다. 그림 4(a)에 이진화된 숫자 영상 샘플을 보였고, 그림 4(b)에 Zhang과 Suen의 알고리즘을 이용한 세선화 결과를 보였다[12]. 그림 4(a)와 같이 잡음이 적은 영상인 경우 세선화된 골격은 문자 영상의 형태를 적절히 유지하면서 획의 폭을 일정하게 해줌을 알 수 있다.

(a) 잡음이 적은 영상

(b) 세선화된 골격

그림 4. 잡음이 적은 문자 영상의 세선화 골격

그러나 세선화는 많은 시간을 요구하는 반복 알고리즘일 뿐만 아니라, 영상에 잡음과 요철이 존재하는 경우 결과되는 골격이 원래 영상의 형태를 적절히 반영하지 못하는 단점을 가진다. 특히 획의 교차점, 분기점, 끝점 등에서 원래 영상의 형태가 왜곡되는 경우가 흔히 발생할 수 있으며, 세선화에 의해 변형된 골격과 파생되는 잔가지를 가지는 영상은 인식 시 오류를 증가시키는 요인으로 작용한다. 실제 응용 분야에서 취득되는 영상은 윤곽선 부근에 다량의 잡음을 포함할 수 있다. 세선화 알고리즘은 윤곽선의

요철이 영상의 일부인지 잡음인지를 분간하기 어려우므로, 세선화 적용 후의 영상에는 요철 부분에 잔가지가 형성되거나 원래 영상의 끝점이 유실되는 등의 영상 왜곡이 일어난다. 이는 근본적으로 획과 요철의 구분이 경험적이고 전역적인 판단에 근거하므로, 지역적 연산에 의존해야하는 세선화 알고리즘 자체가 가지는 문제라고 할 수 있다. Zhang과 Suen의 알고리즘을 비롯해 현재 사용되고 있는 세선화 알고리즘은 이러한 문제점을 어느 정도 완화시킨 것은 사실이나 근본적으로 해결하기에는 미흡하다. 세선화 시에 자주 발생하는 잔가지 문제의 적극적 해결을 위해 Chiu와 Tseng이 제안한 알고리즘은 시간적 복잡도가 높고 영상의 특정 규격을 가정하는 등 일반성이 결여되고, 연산 시간이 오래 걸리더라도 인간의 시각효과를 이용해 그레이 스케일 영상으로부터 직접 세선화를 시도한 경우도 있다[13,14]. 또한 Flores의 알고리즘은 잔가지가 효과적으로 감소하는 대신에 골격이 휘어져 나타나서 원래 형태를 적절히 반영하는데 어려움이 있다[15].

실제 산업현장에서 취득된 영상 중 윤곽선 부근에 잡음을 가지는 영상을 그림 5(a)에 보였다. 그림 5(a)의 영상에 Zhang과 Suen의 세선화 알고리즘을 적용했을 때 결과되는 골격을 그림 5(b)에 보였다. 여기에서 보여지듯이 세선화 알고리즘은 잡음에 매우 민감한 경향을 나타낸다. 잔가지 문제 해결 방법의 하나인 Flores의 알고리즘을 그림 5(a)에 적용한 결과를 그림 5(c)에 보였다. 그림 5(c)에서 그림 5(b)의 잔가지가 일부 제거되었음을 알 수 있으나, 잔가지를 없애는 과정에서 원영상의 골격이 둥글려지거나 분기점이 왜곡되는 등의 부작용을 볼 수 있다. 그림 5(b) 또는 그림 5(c)와 같이 왜곡된 골격을 이용해 구해지는 특징은 인식에 부정적인 영향을 줄 수 있다.

(a) 잡음이 많은 문자 영상

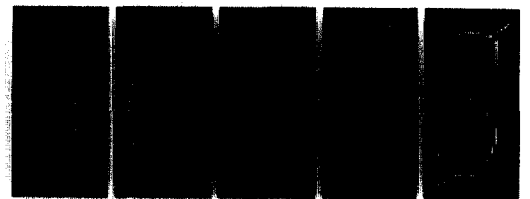
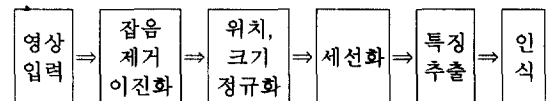
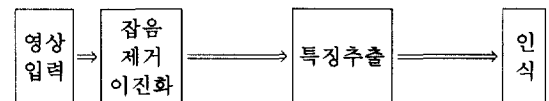
(b) Zhang의 세선화 결과

(c) Flores의 세선화 결과

그림 5. 잡음이 많은 문자 영상의 세선화 골격

(a) 기존의 방법

(b) 제안된 방법

그림 6. 특징 추출 방법

3. 윤곽선 기반 특징

문자 영상을 정규화하거나 세선화하는 영상의 변환 없이 원영상에서 직접 인식에 적합한 특징을 구할 수 있다면, 인식 성능 면에서 보다 견고할 수 있으며 실행 시간 면에서 효율적일 수 있다. 이에 따라 그림 6(a)에서와 같은 기존의 정규화와 세선화 방법을 사용하지 않고, 그림 6(b)와 같이 이진화된 원영상에서 직접 특징을 구하는 방법을 제안한다.

이진화된 문자 영상에서 인식에 적합한 형태적 특징을 신속하게 구하기 위해 다음의 사항들을 고려하였다. 첫째, 문자 영상이 적절히 분할되었고 회전 변화가 없음을 가정하였다. 영상분할과 회전 보상은 특징 추출과 직접적 연관 없이 문서상에서의 문자열, 구획, 종이 위치 등의 전역적 정보를 이용하는 별도의 연구 분야로 다루었다. 둘째, 문자 영상의 크기와 위치 변화를 수용하기 위한 영상 변환을 배제하였다.

위치와 크기 정규화는 추가 연산을 필요로 할뿐만 아니라 원영상의 정보를 유실할 가능성을 가진다. 셋째, 획의 폭을 균일화하기 위한 세선화를 수행하지 않았다. 세선화는 많은 시간을 요하는 반복적 작업이므로 전처리 속도를 저하시킬 뿐만 아니라, 잡음이 많은 경우 세선화 골격이 적절하게 원영상의 형태를 유지하지 못할 수 있기 때문이다.

문자 형태에 대한 정보를 획을 지나는 중간선(medial axe)인 골격에서 구하는 방법이 일반적이나, 문자를 이루는 획의 윤곽선 정보 또한 문자 영상의 형태를 표현하는데 큰 역할을 한다. 따라서 제안된 방법은 문자 영상으로부터 인식에 필요한 특징을 획의 윤곽 부분에서의 방향적 특성에서 구하고자 하였다. 먼저 특정 화소가 획의 윤곽 형성에 기여하는 정도를 나타내기 위하여, 화소 인근에 분포한 배경화소(0인 화소)와 획화소(1인 화소)의 구성비를 사용하였다. 즉, 화소를 중심으로 하는 N X N 창을 설정하고, 화소가 윤곽 형성에 기여하는 정도를 N X N 창 내에서의 배경화소 수와 획화소 수의 비로 나타내었다.

본 논문에서는 획을 구성하는 화소 중에서 획의 윤곽에 인식의 관점에서 의미 있게 분포하는 화소를 윤곽화소로 정의하였다. 획의 윤곽에 어느 정도 가까이 있는 획화소가 인식에 얼마만큼 영향을 주는가를 근사화(approximation)하는 방법으로 획화소(1인 화소) 중에서 화소비가 특정 임계값(임계화소비)을 초과하는 화소를 윤곽화소로 구별하였다. 이 경우 획화소가 인식에 주는 영향은 획화소가 윤곽선 주위에 분포한 정도를 결정하는 임계화소비에 따라 변하게 되므로, 다양한 임계화소비를 이용한 인식 실험을 통해서 임계값 중에서 가장 높은 인식률을 결과하는 값을 최적임계화소비로 사용하였다. 이렇게 구해지는 최적임계화소비를 이용해 윤곽화소를 선택하고 윤곽화소만을 대상으로 방향성분 특징을 구하였다.

영상 격자 안에서 문자 영상의 위치와 크기에 대한 정보를 윤곽화소가 수평축과 수직축에 투영된 양으로 나타내었으며, 수평축과 수직축에 누적된 투영값을 이용하여 영상이 존재하는 실투영 구간을 설정하였다. 여기에서 누적된 투영값은 문자 영상 자체가 투영된 값이 아니라, 획의 윤곽 성분이 투영된 값으로서 윤곽 성분의 위치와 크기 정보를 나타낸다. 문자 영상의 작은 변화가 인식에 미치는 영향을 줄이기 위해 설정된 실투영 구간을 적절한 크기의 등간격으로 분할하였다.

방향연산자를 윤곽화소에 한해 선택적으로 적용하는 것은 획의 윤곽에서만 방향성 정보를 추출함을 뜻한다. 방향 연산자의 한 예로써 그림 7과 같은 4개의 5 X 5 방향(수평, 수직, 사선, 역사선) 마스크 연산자를 사용할 수 있다. 이때 사용되는 방향 마스크의 크기는 윤곽 화소를 정의하는데 사용하는 창과 같은 크기일 필요는 없으며, 사용되는 방향 마스크도 다른 형태의 방향 마스크라도 가능하다. 그림 7과 같은 방향 마스크를 윤곽화소에 적용하여, 획이 가지는 지역적 4방향 성분을 찾을 수 있다. 임의의 숫자 영상에 그림 7의 방향 연산자를 적용한 결과를 그림 8에 보였다.

방향 마스크를 윤곽화소에 선택적으로 적용시킨 결과를 개별 마스크에 할당된 윤곽화소를 투영한 경우와 동일하게 수평축과 수직축에 투영하였다. 개별

0	0	0	0	0
0	0	0	0	0
1	1	1	1	1
0	0	0	0	0
0	0	0	0	0

수평마스크

0	0	1	0	0
0	0	1	0	0
0	0	1	0	0
0	0	1	0	0
0	0	1	0	0

수직 마스크

0	0	0	0	1
0	0	0	1	0
0	0	1	0	0
0	1	0	0	0
1	0	0	0	0

사선 마스크

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

역사선 마스크

그림 7. 5×5 방향 마스크

원영상

수평 마스크

수직 마스크

사선 마스크

역사선 마스크

그림 8. 5×5 방향 마스크 결과

마스크 결과가 투영되는 구간은 윤곽화소가 투영된 실투영 구간과 동일하도록 하였다. 결과적으로, 윤곽화소가 투영된 분할 구간은 개별 마스크 연산 결과가 투영된 분할 구간과 서로 1 : 1 사상된다. 이렇게 서로 사상된 분할 구간에 존재하는 윤곽화소 투영값과 방향특성 투영값의 비를 해당 구간에서의 방향 성분 특징으로 이용하였다.

이러한 투영이 수행되는 과정을 그림 9에 보였다. 그림 9(a)에 원영상과 원영상의 수평축, 수직축 투영 분포를 보였다. 동일한 영상에 대한 윤곽화소의 투영값을 그림 9(b)에 보였다. 윤곽화소에 수평, 수직, 사선, 역사선 마스크를 적용한 영상과 투영값을 그림 9(c,d,e,f)에 각각 보였다. 문자 영상의 특징으로 이용되는 비례값은 그림 9(b)의 윤곽화소 투영값과 그림 9(c,d,e,f)의 방향 특성 투영값 중에서 서로 사상되는 구간의 비례값이다.

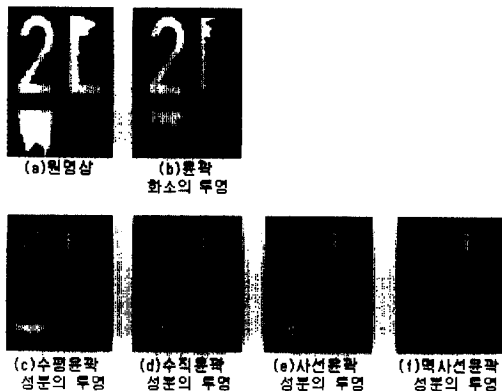



그림 9. 윤곽화소의 방향성분 투영 결과

예를 들어 수평, 수직축을 k개의 등간격으로 나눈다면, 각 마스크 연산은 수평, 수직축에 각각 k개의 비례 값을 생성하여 k차원 벡터를 구성한다. 이러한 k차원 벡터는 각 마스크마다 쌍으로 만들어져, 총 8개의 k차원 벡터가 구성된다. 각 k차원 벡터의 성분은 원영상의 동일 투영 구간에서 윤곽화소 투영값과 방향 성분 투영값의 비를 뜻하므로, 영상의 위치와 크기 변화를 수용함과 동시에 획의 두께에 무관한 방향 성분 특성을 가지게 된다. 따라서 전체 8개의 k차원 벡터는 문자 영상의 위치와 크기의 변화를 수용하면서 문자 영상을 구성하는 획의 방향 성분 모두를 포함하게 되어, 문자 영상의 전역적 특징을 효과적으로 나타낼 수 있게 된다. 문자 영상으로부터 위치, 크기, 획 폭의 변화를 수용하는 윤곽의 전역적 방향 특성을 추출하는 단계를 아래에 요약하였다.

단계1) 이진 영상을 주사하면서 각 화소에 N X N 창 내의 배경화소(0인 화소) 수와 획화소(1인 화소) 수의 비를 구한다.

단계2) 획화소 중에서 임계화소비를 초과하는 윤곽화소에 한하여,
- 윤곽화소를 수평축, 수직축에 투영한다.
- 윤곽화소에 대해 4방향 마스크 연산 결과를 각 마스크에 할당된 수평축, 수직축에 투영한다.

단계3) 주사가 완료되면,
- 윤곽화소가 투영된 수평축, 수직축에서 윤곽화소의 실투영 구간을 설정한다.
- 윤곽화소 실투영 구간을 최적 등분 분할한다.
- 각 마스크에 할당된 수평축, 수직축을 윤곽화소가 투영된 수평축, 수직축에 1:1 사상한다.
- 사상된 분할 구간에 존재하는 투영값의 비를 구한다.

단계4) 단계3)에서 구한 투영값의 비로 특징 벡터를 구성한다.

제안된 방법은 위치, 크기 정규화와 세선화 등의 별도의 전처리 과정을 필요로 하지 않을 뿐만 아니라, 한번의 영상 주사로 특징 추출 과정이 완료된다. 더욱이 영상의 배경과 영상의 내부에서는 마스크 연산이 수행되지 않으므로, 특징 추출을 위한 전처리 과정에 소요되는 시간을 크게 줄이고 인식의 효율을 높이는 특징을 구할 수 있다.

4. 실험 결과 및 분석

제안된 방법의 인식 효과를 실험하기 위하여 50 X 50 크기의 필기체 자음 및 모음 20,000개 집단과 30 X 60 크기의 자동차 번호판 숫자 512개 집단을 이진화하여 사용하였다. 윤곽화소를 정의하는 창의 크기는 문자 영상의 크기에 따라 정해질 수 있으며 본 실험에서는 5 X 5 창을 사용하였다. 방향 마스크로는 윤곽에 분포한 잡음에 덜 민감하면서 전체적인 획의 방향 특성 추출에 유리한 그림 7의 5 X 5 마스

크 연산자를 사용하였다.

다양한 방법에 의해 구해지는 특징들간의 인식성능 비교는 신경회로망의 구조에 독립적인 동일한 조건에서 비교되어야 하므로, 모든 인식 실험에서 다음과 같은 동일한 인식 시스템을 사용하였다. 하나의 은닉층을 가진 다층 퍼셉트론을 사용하였으며, 은닉층 노드의 수는 16개, 가중치 교정율은 0.1로 통일하여 학습에 사용하였다. 입력층과 출력층 노드 수는 사용되는 특징과 집단의 종류에 따라 가변되며, 학습에 소요되는 시간도 이러한 조건에 종속된다. 이에 따라 모든 실험에서 신경회로망이 충분히 수렴될 수 있도록 시간적 제약을 완화하였다.

먼저 필기체 자모음 문자의 인식을 실험하였으며, 이 경우 수평, 수직축의 실투영 구간은 3등분으로 분할하였다. 하나의 마스크 연산은 수평축과 수직축에 각각 3차원 벡터를 결과하며, 4방향 마스크 연산으로 총 24개의 실수로 구성되는 특징 벡터를 얻게 된다. 주어진 창에서 0인 배경화소 수가 n개이고 1인 획화소 수가 m개인 경우, 화소비는 n/m으로 나타낼 수 있다. 따라서 5 X 5 창에서는 화소비가 1/24부터 12/13까지의 값을 가질 수 있다. 윤곽화소를 정의하는 효과적인 임계화소비를 결정하기 위하여, 임계화소비를 1/24부터 단계적으로 증가시키면서 인식성능을 측정하였다. 각 임계화소비를 이용해 구해진 패턴 집단을 다층 퍼셉트론으로 동일한 시간 학습시킨 후, 인식 결과를 출력층에서 측정된 인식오류율과 평균제곱오류율로 표 1에 보였다. 인식오류률은 임계화소비가 증가함에 따라 감소하다가 다시 증가하는 함수 형태로 나타났으며, 이러한 결과는 획의 윤곽에 매우 근접하거나 또는 획으로부터 멀리 있는 획화소는 획의 윤곽 특성을 상대적으로 적게 가짐으로서 인식에 적은 영향을 주는데 기인한다. 임계화소비가 3/22일 때 가장 낮은 인식오류률을 보임에 따라 최적

표 1. 임계화소비에 따른 인식 결과(%)

임계화소비	인식오류율	평균오류율
1/24	4.37	0.35
2/23	3.96	0.33
3/22	2.84	0.25
4/21	4.48	0.36
5/20	6.64	0.57
6/19	11.25	0.95

임계화소비로 결정하였으며, 본 실험에서는 획화소중에서 최적임계화소비 3/22를 초과하는 화소에 한하여 방향 마스크 연산을 적용하였다.

실투영 구간의 등분 수와 인식 성능의 관계를 알아보기 위하여, 수평, 수직축 실투영 구간을 2등분, 3등분, 4등분 분할한 각각의 경우에 대한 인식 실험을 수행하였다. 표 2의 인식 결과에 의하면 2등분 분할인 경우 16차원 특징 벡터를 구성하게 되며, 인식성능이 매우 낮게 나타났다. 3등분과 4등분 분할인 경우 각각 24차원 특징 벡터와 32차원 특징 벡터를 구성하였다. 표 2의 결과는 2등분 분할에 비해 3등분 분할이 크게 향상된 인식률을 보이는 반면에, 3등분 분할과 4등분 분할에서는 큰 차이를 나타내지 않음을 보여준다. 특히 4등분 분할의 경우 특징 벡터의 크기가 33% 이상 증가하여 학습과 인식에 더 많은 시간을 필요로 하게되는데 반하여 오히려 인식오류가 다소 증가함을 보였다. 이로써 분할 구간을 더 늘려도 그에 수반되는 인식 효과가 크지 않음을 예상할 수 있으므로, 본 실험에서는 수평, 수직축의 실투영 구간을 3등분 분할하였다.

표 2. 실투영구간의 등분 수에 따른 인식 결과(%)

분할 수	특징벡터 크기	인식오류율	평균오류율
2등분	16	10.21	0.84
3등분	24	2.84	0.25
4등분	32	3.14	0.27

정규화 센트랄모우먼트(normalized central moment)는 제안한 방법과 동일하게 문자영상의 크기, 위치, 획의 폭을 정규화 하지 않고 특징을 추출하는 방법이다. 따라서 제안된 방법과 인식 성능과 속도를 비교하고자, 정규화 센트랄모우먼트를 이용하여 비교 인식 실험을 하였다. N X M 크기의 문자 영상이 불연속 함수 $f(i,j)$로 표시되면, 차수가 (p,q)인 2차원 모우먼트(moment)는

$$m_{pq}=\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} i^p j^q f(i,j), \qquad p,q=0,1,2,\ \ldots\ldots,$$

로 나타내어진다. 또한 센트랄모우먼트(central moment)는 $i_c=m_{10}/m_{00}$, $j_c=m_{01}/m_{00}$인 경우,

$$\mu_{pq}=\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(i-i_c)^p\ (j-j_c)^q f(i,j)$$

로 표시된다. 위의 식으로 구해진 센트랄모우먼트는 영상의 위치 변화에 무관하게 된다. 영상의 크기에 독립적인 정규화 센트랄모우먼트(normalized central moment)는 $\eta_{pq} = \mu_{pq} / \mu_{00}{}^{r}$로 주어진다. 여기서 $r = (p+q)/2 + 2$이다[16].

이 실험에서는 영상의 작은 변화에 덜 민감한 8개의 낮은 차수의 정규화 센트랄모우먼트를 특징으로 이용하였다. 제안된 방법과 정규화 센트랄모우먼트에 의한 인식 결과를 표 3에 보였다. 표 3에서 정규화 센트랄모우먼트를 이용한 인식 결과가 제안된 방법에 비해 높은 인식오류율인 5.03%를 보여, 제안된 방법이 인식 성능 면에서 효과적임을 나타냈다. 문자 영상에서 특징을 구하는데 요구되는 시간적 복잡도를 비교하고자, 제안한 방법에 소요되는 시간과 정규화 센트랄모우먼트를 구하는데 소요되는 시간을 PC Pentium Pro 200에서 실행하여 표 3에 보였다. 제안된 방법은 등분 분할의 수에 무관하게 일정한 시간이 소요됨을 알 수 있다. 이러한 결과는 제안된 방법이 등분 분할의 수와 관계없이 영상을 한 번만 주사하여 완료되는데 기인한다. 반면에 적은 수의 특징인 8개의 정규화 센트랄모우먼트를 구하는데 소요된 시간은 제안된 방법보다 약 3배 이상의 시간이 필요함을 보여 제안된 방법이 소요 시간 면에서 보다 효율적임을 보였다.

표 3. 제안된 방법과 정규화 센트랄모우먼트의 인식 비교

특징 추출 방법	인식오류률 (%)	평균오류률 (%)	소요시간 (msec)
제안방법 (2등분 분할)	10.21	0.84	7.53
제안방법 (3등분 분할)	2.84	0.25	7.59
제안방법 (4등분 분할)	3.14	0.27	7.88
정규화 센트랄모우먼트	5.03	0.60	24.41

문자 인식에 있어서 일반적으로 가장 많이 사용하는 방법은 위치와 크기의 변화를 먼저 정규화한 후 세선화에 의해 골격 영상을 구성한 다음에 특징을 구하는 것이다. 이에 따라 필기체 문자 집단에 대해 다음과 같은 비교 실험을 하였다. 첫 번째는 위치와 크기가 정규화되지 않은 영상에서 세선화한 골격의 망 특징을 구하였다. 두 번째는 위치와 크기를 정규화한 후 세선화한 골격에서 망 특징을 구하였다. 세 번째로는 두 번째의 상태에서 방향 마스크를 적용한 후 망 특징을 구하였다. 이상의 3가지 특징에 의한 인식 결과를 표 4에 요약하였다. 문자 영상의 세선화 알고리즘으로는 보편적으로 많이 이용되고 있는 Zhang과 Suen이 제안한 알고리즘을 사용하였다.

표 4. 제안된 방법과 정규화, 세선화 방법의 인식 비교

특징 추출 방법	인식오류률 (%)	평균오류률 (%)	특징벡터 크기
제안방법 (3등분 분할)	2.84	0.25	24
세선화+망 특징	36.43	2.57	25
위치,크기정규화 +세선화+ 망 특징	7.21	0.59	25
위치,크기정규화 +세선화+방향 마스크+망 특징	7.51	0.63	100

먼저 첫 번째 방법은 예상대로 매우 나쁜 인식 결과로 나타났다. 이는 위치와 크기의 정규화가 망 특징에 필수적임을 보여준다. 두 번째 방법은 위치, 크기의 정규화와 세선화 후에 10 X 10 크기로 압축한 망 특징을 사용한 경우로서 인식오류률이 7.21%로 나타났다. 세 번째 방법에서는 4방향 성분의 망 특징을 사용함에 따라 특징 벡터의 크기가 두 번째 방법에 비해 4배로 증가했다. 그럼에도 불구하고 인식 성능은 큰 차이가 없게 나타났다. 3가지 방법 모두 제안된 방법에 비해 인식 성능이 떨어짐을 보였다. 이는 원영상에 정규화나 세선화를 적용할 경우 원영상의 정보가 유실되거나 왜곡될 수 있으며, 결과적으로 인식에 부정적인 영향을 줄 수 있음에 그 원인이 있다고 판단된다. 특히 잡음이 많은 영상에 세선화 알고리즘을 적용할 경우에 이러한 현상이 발생할 가능성이 증가한다. 특징 추출에 소요되는 시간은 정규화와 세선화를 적용하는 방법이 제안된 방법보다 2.5배 내지 3배정도 더 소요되는 것으로 나타났다. 필기체 문자 인식 실험에 사용된 문자 영상 중에서 오인식을 유발한 영상의 예를 제안된 방법, 정규화 센트랄모우

먼트 방법, 위치, 크기 정규화 후 세선화, 망 특징 방법의 경우로 구분하여 그림 10에 보였다.

(a)

(b)

ㅠ⇒ㅈ ㄹ⇒ㅎ ㅅ⇒ㄴ ㅈ⇒ㅠ ㅁ⇒ㅇ

(c)

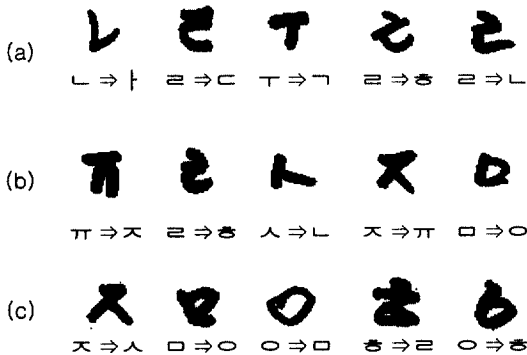



그림 10. 필기체 문자 중 오인식된 영상

(a) 제안된방법, (b) 정규화센트랄모우먼트, (c) 위치,크기 정규화+세선화+망 특징

다음으로는 CCD 카메라로 자동차 번호판 영상을 입력한 후, 번호판 영상에서 큰 숫자를 분할한 영상에 대한 인식 실험을 수행하였다. 이때 영상의 크기는 30 X 60이고, 총 516개의 영상 중에서 비교적 상태가 양호한 250개의 영상을 이용해 신경회로망을 학습시켰다. 나머지 266개의 영상은 테스트 패턴으로 사용하였다. 제안된 방법은 수직축을 6등분, 수평축을 3등분하여 36차원 특징벡터를 구성하였다. 비교방법으로 위치와 크기 정규화, 세선화, 방향마스크, 망 특징에 의해 72차원 특징벡터를 구성하였다. 표 5에 두 가지 방법에 의한 인식 결과를 비교하였다. 학습 패턴에 대해 두 방법 모두 완전하게 인식하였으나, 테스트 패턴의 인식에서는 제안된 방법이 약 4% 정도 향상된 인식률을 보였다. 두가지 방법에 의한

표 5. 자동차 번호판 숫자 인식 결과

특징 추출 방법	특징 벡터 크기	학습데이터 인식오류률 (%)	테스트데이터 인식오류률 (%)
제안방법 (6,3등분분할)	36	0.00	0.93
위치,크기 정규화+세선화+방향 마스크+망 특징	72	0.00	13.16

자동차 번호 인식 실험의 경우 인식에 실패하는 영상들은 그림 11과 같다.

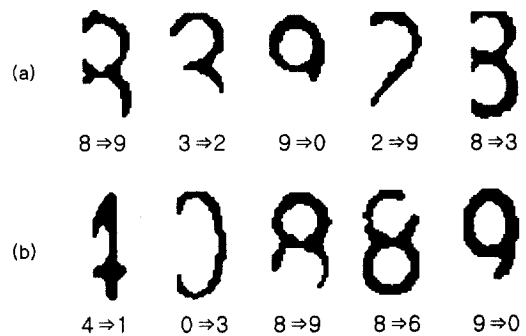



그림 11. 자동차 번호 중 오인식된 영상

(a) 제안된 방법,
(b) 위치,크기 정규화+세선화+망 특징

다층 퍼셉트론은 일반화 특성의 장점이 있는 반면에 복잡한 사상을 요구하는 문자 인식에서 극소해 귀착 가능성이 커질 수 있다. 국소적 학습 방식의 다층 퍼셉트론의 극소해 가능성을 줄이고자, 전역적 최적화를 수행하는 유전자 알고리즘을 이용해 가중치를 초기화하였다[17,18]. 가중치 초기화 방법은 가중치를 병렬시킨 스트링을 구성하여 하나의 개체로 정의하여, 개체의 적합도를 패턴 집단에 대해 순방향 전파를 수행한 평균 출력 오류값으로 하였다. 인식 실험 결과 유전자 알고리즘으로 초기화된 가중치에 의한 학습은 무작위 가중치 할당의 경우보다 수렴에 필요한 반복 횟수는 줄어들었으나, 초기화하는데 보다 많은 시간이 소요되므로 학습 시간 측면에서 유리하다고 하기는 어려웠다. 그러나 최적화된 초기 상태로부터의 학습은 극소해에 귀착하는 가능성을 줄일 수 있으므로 시행착오적인 반복 실험 횟수를 줄이는데 도움이 되었다.

5. 결 론

문자 인식을 위한 신경회로망의 학습은 고속 컴퓨터나 병렬, 분산 컴퓨팅 환경에서 장시간 오프라인으로 학습시킬 수 있으므로 시간적 제약이 완화될 수 있다. 그러나 인식에 필요한 문자 영상의 전처리는

실시간 온라인으로 실행되므로, 인식에 적합한 특징을 신속하게 추출해야 한다. 잡음을 포함하거나 작은 문자 영상에 대한 기하학적 정규화나 세선화 변환은 원영상의 정보를 유실하거나 왜곡하여 인식에 부정적인 효과를 줄 수 있다. 특히 세선화는 잡음이나 세부 형태의 변화에 매우 민감하므로, 세선화된 골격이 문자 영상의 원형을 적절히 유지하지 못할 경우 인식에 부정적인 결과를 줄 수 있다.

이에 따라 본 연구는 오프라인 문자인식시스템의 구현에 있어, 문자 영상의 인식에 필요한 효과적인 특징을 신속하게 원영상에서 직접 구하고자 하였다. 문자 영상의 주요 특징을 구하기 위한 방법으로 획을 구성하는 화소 중에서 인식의 관점에서 의미 있는 화소를 윤곽화소로 정의하고, 이러한 윤곽화소에 한하여 선택적으로 방향 성분을 구하고자 하였다. 윤곽화소에 마스크 연산을 적용한 방향 성분을 영상 축에 투영하였으며, 투영된 구간을 최적 균등 분할하여 서로 사상되는 구간에 대한 투영값의 비를 특징으로 사용하였다. 문자 영상의 위치, 크기, 획의 폭 변화를 원영상의 변환 없이 수용하므로 인식에 효과적인 특징을 구할 수 있었으며, 한 번의 영상 주사로 모든 과정이 이루어지게 함으로써 전처리에 소요되는 시간을 줄일 수 있었다.

자모음 인식 실험에서 정규화 센트랄모우먼트 방법과 위치, 크기 정규화 후 세선화한 방법은 각각 인식오류율 5.03%와 7.21%를 보인 반면에, 제안된 방법은 인식오류율 2.84%를 나타내 인식에 보다 효과적임을 보였으며, 실행시간 면에서는 제안된 방법이 3배정도 효율적이었다. 자동차 번호 인식의 경우에서도 제안된 방법이 정규화 후 세선화 방법보다 4% 정도 우수한 인식률을 보였다. 이러한 인식 성능의 차이는 크기가 작고 잡음이 많은 문자 영상을 정규화하고 세선화하는 영상 변환이 인식에 부정적으로 작용할 수 있음에 반하여, 제안된 방법은 영상 변환 없이 직접 원영상으로부터 효과적인 특징을 구하는데 기인한다고 판단된다. 또한 제안된 방법은 간단한 마스크 연산과 한번의 화상 주사로 특징 추출이 완료되므로 인식 속도 면에서 유리하였다. 제안된 방법이 인식에 효과적인 특징을 신속하게 구할 수 있으므로 산업 현장의 실시간 문자 입력 자동화에 도움이 되리라 기대한다.

참 고 문 헌

[1] C. Y. Suen, M. Berthod, and S. Mori, “Automatic Recognition of Handprinted Character - The State of the Art,” Proceedings of the IEEE, Vol.68, No.4, 1980.

[2] 이희건, 방승양, “신경회로망을 이용한 오프라인 문자인식 연구 동향,” 정보과학회지 제9권, 제1호, pp. 33-45, 1991.

[3] S. Liang, M. Shridhar, and M. Ahmadi, “Segmentation of Touching Characters in Printed Document Recognition,” Pattern Recognition, Vol.27, No.6, pp. 825-840, 1994.

[4] R. M. Brown, T. H. Fay and C. L. Walker, “Handprinted Symbol Recognition System,” Pattern Recognition, Vol.21, No.2, pp. 91-118, 1988.

[5] L. Lam, S. Lee, “Thinning Methodologies - A Comprehensive Survey,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.14, No.9, pp. 869-885, Sep. 1992.

[6] C. J. Hilditch, “Linear Skeletons from Square Cupboards,” Machine Intelligence, American Elsevier, New York, USA, Vol.4, pp. 403-420, 1969.

[7] D. E. Rumelhart, J. L. McClelland and the PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol.1, M.I.T. Press, 1986.

[8] 김우태, 윤병식, 박인규, 진성일, “인쇄체 한글 문자인식을 위한 특징성능의 비교,” 한국정보과학회 논문지 제20권, 제8호, pp. 1103-1111, 1993.

[9] 김영준, 이성환, “유전자 알고리즘과 결합된 다층 클러스터 신경망을 이용한 무제약 필기체 숫자의 오프라인 인식,” 한국정보과학회 논문지 제21권, 제8호, pp. 1468-1479, 1994.

[10] 박창순, 김두영, “오프라인 필기체 숫자 인식을 위한 다양한 특징들의 성능 비교 및 인식률 개선 방향,” 한국정보처리학회 논문지 제3권, 제4호, pp. 915-925, 1996.

[11] C. Teh and R. T. Chin, "On Digital Approximation of Moment Invariant," Computer Vision, Graphics, and Image Processing, Vol.33, pp. 318-326, 1986.

[12] T. Y. Zhang and C. Y. Suen, "A Fast Algorithm for Thinning Digital Pattern," Comm. ACM. Vol.25, No.3, pp. 236-239, 1984.

[13] H. Chiu and D. Tseng, "A Feature-Preserved Thinning Algorithm for Handwritten Chinese Characters," Proceedings of the ICPR '96, pp. 235-239, 1996.

[14] Y. Chen and Y Yu, "Thinning Noisy Binary Pattern Using Human Visual Symmetry," Proceedings of ICPR '96, pp. 146-150, 1996.

[15] E. Flores, "A Fast Thinning Algorithm," Proceedings of the SBT/IEEE International Telecommunications Symposium, Vol.2, pp. 594-599, 1998.

[16] M. Hu, "Visual Pattern Recognition by Moment Invariants," IRE Trans. on Information Theory, Vol.8, pp. 179-187, 1962.

[17] J. H. Holland, Adaptation in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor, M.I., 1975

[18] D. E. Goldberg, "Genetic and evolutionary algorithms come of age," Communications of the ACM, Vol.39, No.3, pp. 113-119, 1994.

공 용 해

1982년 연세대학교 전자공학과 졸업(공학사)
1986년 미국 폴리텍 대학원 전산학과 졸업(전산학석사)
1991년 미국 폴리텍 대학원 전산학과 졸업(전산학박사)
1982년 한진중공업 연구원
1983년 삼성전자 연구원
1991년~현재 순천향대학교 컴퓨터학부 부교수
관심분야 : 신경회로망, 패턴인식, 컴퓨터비전, 에이전트 등